



UNIVERSIDAD  
**COMPLUTENSE**  
MADRID

Proyecto de Innovación

Convocatoria 2016/2017

Nº de proyecto 266

Título del proyecto:

Integración de software estadístico R en el entorno Moodle: Elaboración de casos prácticos de Estadística Descriptiva

Nombre del responsable del proyecto:

Francisco Javier Martín Campo

Facultad de Ciencias Matemáticas

Departamento de Estadística e Investigación Operativa

**Contenido**

Objetivos propuestos en la presentación del proyecto .....3

Objetivos alcanzados .....4

Metodología empleada en el proyecto .....5

Recursos humanos.....6

Desarrollo de las actividades .....7

Anexo .....8

## Objetivos propuestos en la presentación del proyecto

En la solicitud del proyecto se plantearon los siguientes objetivos, clasificados en generales y específicos.

### **Objetivos generales:**

1. Fomentar el aprendizaje autónomo del alumno en materias de estadística como complemento a la formación recibida en el aula.
2. Fomentar un ritmo de trabajo regular basado en la realización de ejercicios

### **Objetivos específicos:**

1. Promover la participación del alumno en la plataforma Moodle.
2. Promover el autoaprendizaje del alumno mediante contenidos motivadores.
3. Promover el trabajo colectivo de los alumnos mediante casos prácticos más elaborados que exijan la participación colectiva del alumnado.
4. Incrementar la motivación del alumno en las asignaturas de carácter estadístico.
5. Implementar los casos prácticos usando software estadístico R.

## Objetivos alcanzados

Durante la realización del proyecto, se han implementando casos de estudio que formarán parte de una base de datos de problemas más general. Los casos de estudio han sido implementados usando el software estadístico R, que mediante una librería determinada permite exportar un documento xml propio de Moodle para su integración en esa plataforma.

La temática abordada en los problemas implementados es estadística descriptiva. Estos problemas estarán dirigidos a alumnos de las Facultades de Informática y Farmacia principalmente, donde nuestro departamento imparte docencia. Una importante parte de los alumnos de las titulaciones en estas facultades han mostrado tener dificultades en superar materias de contenido matemático y estadístico. Es por ello, que esta base de problemas que se está desarrollando es importante ya que permitirá al alumno autoevaluarse desde su propia casa.

Un problema diseñado bajo la librería *exams* de R tiene la particularidad de que, a pesar de tratarse del mismo problema, cada alumno recibe datos diferentes. Por otro lado, una vez que el alumno termina su actividad, puede automáticamente saber en qué preguntas ha acertado o fallado. Finalmente, el alumno puede ver cómo hubiera sido la resolución correcta del problema. Todo ello favorece el aprendizaje autónomo del alumno.

La implementación de los problemas ha sido laboriosa, y a pesar de haber construido unos pocos problemas, se desea continuar con la elaboración de más casos para enriquecer la base de problemas.

## Metodología empleada en el proyecto

La metodología empleada en el proyecto ha estado basada en las siguientes fases:

1. Recopilación de problemas basados en estadística descriptiva.
2. Implementación de problemas usando software estadístico R.
3. Validación de los problemas creados.

Los casos prácticos han sido creados por profesores que imparten asignaturas con contenido de estadística descriptiva en los Grados de Ingeniería Informática, Ingeniería del Software, Ingeniería de Computadores, Desarrollo de Videojuegos (impartidos en la Facultad de Informática), el Grado en Farmacia (impartido en la Facultad de Farmacia) y el Grado en Administración y Dirección de Empresas (impartido en la Facultad de CC. Económicas y Empresariales).

El trabajo conjunto de los casos prácticos en las fases anteriormente presentadas desemboca en la integración como actividad evaluable dentro de la plataforma Moodle, con el principal objetivo de explotar al máximo las capacidades de la misma en cuanto a calificación se refiere.

El desarrollo del trabajo se ha estructurado en los siguientes niveles:

1. Definición de los problemas prácticos a incluir en la base de problemas.
2. Apertura de un seminario de trabajo con acceso a todos los profesores del equipo, para dar soporte a los casos prácticos.
3. Elaboración de casos prácticos enfocados a estadística descriptiva.
6. Integración de los casos prácticos en Moodle.

Todo ello se ha elaborado en un espacio de trabajo dentro de la plataforma Moodle titulado "PIMCD 266", cuyo acceso es restringido a los autores del proyecto, pero previa petición, puede ser explorado por cualquier interesado/a. En el Anexo puede encontrarse un caso de estudio detallado.

## Recursos humanos

Los recursos humanos con los que ha contado el proyecto han sido los profesores integrantes del equipo formado:

- Rosa Alonso Sanz
- F. Javier Martín Campo
- M. Teresa Ortuño Sánchez
- J. Tinguaro Rodríguez González
- Karina H. Rojas Patuelli
- Gregorio Tirado Domínguez

Se ha solicitado un nuevo proyecto como continuación de este con la intención de incrementar la base de problemas. El proyecto ha sido también concedido y se espera que en una fase posterior a la finalización del nuevo proyecto se pongan a disposición de los alumnos para estudiar su impacto en sus resultados académicos.

## Desarrollo de las actividades

Para la consecución de los objetivos descritos en la propuesta, se han planteado las siguientes actividades:

1. Estudio del artículo de referencia en nuestro proyecto: “*Flexible generation of e-learning exams in R: Moodle quizzes, OLAT assessments, and beyond*”, de A. Zeileis, N. Umlauf y F. Leisch en *Journal of Statistical Software* 58(1), pp.: 1-36 y creación de los primeros casos prácticos.
2. Reunión con el equipo del proyecto donde se expusieron las dificultades adquiridas en la implementación de los primeros problemas.
3. Implementación de casos prácticos relacionados con estadística descriptiva.
4. Validación de los casos prácticos implementados.
5. Incorporación de enunciados a los casos implementados.

## Anexo

A continuación se presentan capturas de pantalla de un caso práctico creado, con el fin de ilustrar el material creado por el equipo de trabajo.

El seminario virtual de trabajo presenta la siguiente estructura:

The screenshot shows the Moodle interface for the course 'seminario-invest-59704-5'. The top header displays 'CVUCM-Moodle29' and the language 'Español - Internacional (es)'. The main content area is titled 'PIMCD 266'. On the left, a 'NAVEGACIÓN' sidebar lists the course structure: 'Mi Campus', 'Área personal', 'Páginas del sitio', 'Curso actual', 'seminario-invest-59704-5' (expanded), 'Participantes', 'Insignias', 'General', 'Ejemplos', 'Nuevos (grupo)', 'Mi Correo', and 'Mis cursos'. The main content area is divided into two sections: 'General' and 'Ejemplos'. The 'General' section includes a 'Novedades' icon. The 'Ejemplos' section lists three items: 'Artículo R', 'Ejemplo con tres preguntas', and 'Contraste de hipótesis para la diferencia de medias', each with a document icon and a checkmark.

Se cuenta con una serie de ejemplos creados por previamente que sirvieron de punto de partida para la creación de los nuevos.

Cada cuestionario creado puede constar del número de problemas deseado, donde las diferencias están en los datos y no en la descripción del problema. De este modo, el profesor garantiza que todos los alumnos resolverán el mismo problema, pero con datos distintos.

A modo de ejemplo, se pretende ilustrar el funcionamiento de los mismos con capturas de pantalla. Primeramente el alumno accede a un cuestionario y es remitido al enunciado del mismo:



## Vista previa de la pregunta

### Pregunta 1

Sin responder aún

Puntúa como 41,00

En un grupo de 100 alumnos de la Universidad Complutense de Madrid, se estudia el número de asignaturas aprobadas en el primer cuatrimestre ( $X$ ) y el número de horas semanales ( $Y$ ).

1. Completar la tabla marginal de frecuencias de la variable  $X$ . A partir de la información dada en la tabla original, se pueden deducir el resto de valores, quedando la tabla de frecuencias:

$X$	$n_i$	$N_i$	$f_i$	$F_i$
0	<input type="text"/>	7	<input type="text"/>	<input type="text"/>
1	23	<input type="text"/>	<input type="text"/>	<input type="text"/>
2	<input type="text"/>	<input type="text"/>	0.23	<input type="text"/>
3	<input type="text"/>	71	<input type="text"/>	<input type="text"/>
4	<input type="text"/>	<input type="text"/>	<input type="text"/>	0.93
5	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>

2. Completar la tabla marginal de frecuencias de la variable  $Y$ . A partir de la información dada en la tabla original, se pueden deducir el resto de valores, quedando la tabla de frecuencias:

$Y$	$n_i$	$N_i$	$f_i$	$F_i$
0-10	<input type="text"/>	<input type="text"/>	0.2	<input type="text"/>
10-20	30	<input type="text"/>	<input type="text"/>	<input type="text"/>
20-30	<input type="text"/>	80	<input type="text"/>	<input type="text"/>
30-40	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>

3. El número máximo de asignaturas que tiene que haber aprobado un alumno para estar entre 73% de alumnos que menos asignaturas han aprobado es:

El número de horas de estudio a la semana que tiene que dedicar un alumno como mínimo para estar entre el 35% de los que más horas ha estudiado es:

4. El número de asignaturas aprobadas más habitual es:  (nota: en el caso de obtener más de un valor indicar sólo el más pequeño)

Teniendo en cuenta que la distribución de los datos es uniforme dentro de cada intervalo, el número de horas semanales dedicadas al estudio más frecuente es:  (nota: en el caso de obtener más de un valor indicar sólo el más pequeño)

5. Sabiendo la siguiente información de las variables  $X$  e  $Y$ ,

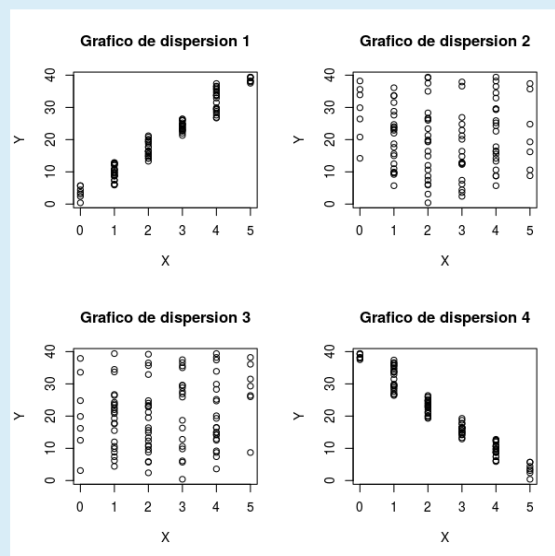
$$\sum_{i=1}^{100} x_i = 246 \quad \sum_{i=1}^{100} x_i^2 = 804$$

$$\sum_{i=1}^{100} y_i = 2057.2 \quad \sum_{i=1}^{100} y_i^2 = 53343.78$$

El coeficiente de correlación lineal es:

El coeficiente de determinación es:

De las siguientes figuras:



La que mejor se ajusta a los resultados obtenidos es el diagrama de dispersión número:

6.

7. El número esperado de asignaturas que aprobaría un alumno que estudia 32 horas a la semana es:

Como se puede contemplar, el alumno tiene distintas opciones de respuesta. En este caso de estudio, menús desplegables e introducción de valores numéricos. También son aceptadas otro tipo de respuestas como múltiples respuestas.

El alumno por tanto deberá responder a las preguntas planteadas:

### Vista previa de la pregunta

**Pregunta 1**  
Sin responder aún  
Puntúa como 41,00

En un grupo de 100 alumnos de la Universidad Complutense de Madrid, se estudia el número de asignaturas aprobadas en el primer cuatrimestre ( $X$ ) y el número de horas semanales ( $Y$ ).

1. Completar la tabla marginal de frecuencias de la variable  $X$ . A partir de la información dada en la tabla original, se pueden deducir el resto de valores, quedando la tabla de frecuencias:

$X$	$n_i$	$N_i$	$f_i$	$F_i$
0	7	7	0.07	0.07
1	23	30	0.23	0.3
2	23	53	0.23	0.53
3	18	71	0.18	0.71
4	22	93	0.22	0.93
5	7	100	0.07	1

2. Completar la tabla marginal de frecuencias de la variable  $Y$ . A partir de la información dada en la tabla original, se pueden deducir el resto de valores, quedando la tabla de frecuencias:

$Y$	$n_i$	$N_i$	$f_i$	$F_i$
0-10	20	20	0.2	0.2
10-20	30	50	0.3	0.5
20-30	30	80	0.3	0.9
30-40	19	97	0.17	0.98

3. El número máximo de asignaturas que tiene que haber aprobado un alumno para estar entre 73% de alumnos que menos asignaturas han aprobado es:

3

El estadístico de contraste es  $a$  y su distribución de probabilidad Normal Estándar

2. El contraste que se desea realizar es:

$H_0$ :  $\mu = 120,28$

$H_1$ :  $\mu \neq 120,28$

3. El valor del estadístico de contraste es: -0,17

4. De las siguientes opciones:

(a)  $R_c = \{|z_c| \geq |z_{\alpha/2}|\}$

(b)  $R_c = \{z_c \geq z_{\alpha}\}$

(c)  $R_c = \{z_c \leq z_{\alpha}\}$

(d)  $R_c = \{z_c \geq -z_{\alpha}\}$

(e)  $R_c = \{z_c \leq -z_{\alpha}\}$

la región crítica es la opción a

5. En base a los resultados obtenidos, hay evidencia estadística para Rechazar la hipótesis nula.

6. El p-valor de la muestra es: 0,05

Siguiente

Proyecto PIMCD 2015 - 133

Una vez que el alumno termina el cuestionario, pulsando en el botón “siguiente” y aceptando el envío del cuestionario, el alumno puede comprobar la puntuación obtenida, sus respuestas y si éstas eran correctas o no.

## Vista previa de la pregunta

### Pregunta 1

Parcialmente  
correcta

Puntuación como 41,00

En un grupo de 100 alumnos de la Universidad Complutense de Madrid, se estudia el número de asignaturas aprobadas en el primer cuatrimestre ( $X$ ) y el número de horas semanales ( $Y$ ).

1. Completar la tabla marginal de frecuencias de la variable  $X$ . A partir de la información dada en la tabla original, se pueden deducir el resto de valores, quedando la tabla de frecuencias:

$X$	$n_i$	$N_i$	$f_i$	$F_i$
0	7	7	0.07	0.07
1	23	30	0.23	0.3
2	23	53	0.23	0.53
3	18	71	0.18	0.71
4	22	93	0.22	0.93
5	7	100	0.07	1

2. Completar la tabla marginal de frecuencias de la variable  $Y$ . A partir de la información dada en la tabla original, se pueden deducir el resto de valores, quedando la tabla de frecuencias:

$Y$	$n_i$	$N_i$	$f_i$	$F_i$
0-10	20	20	0.2	0.2
10-20	30	50	0.3	0.5
20-30	30	80	0.3	0.9
30-40	19	97	0.17	0.98

3. El número máximo de asignaturas que tiene que haber aprobado un alumno para estar entre 73% de alumnos que menos asignaturas han aprobado es:

3

Una vez vistas las respuestas contestadas correcta e incorrectamente, el alumno tiene acceso a una descripción de la solución.

En un grupo de 100 alumnos de la Universidad Complutense de Madrid, se estudia el número de asignaturas aprobadas en el primer cuatrimestre ( $X$ ) y el número de horas semanales ( $Y$ ).

1. Completar la tabla marginal de frecuencias de la variable  $X$ .

Con la información del problema es fácil determinar el resto de valores de la tabla de frecuencias sin más que tener en cuenta las definiciones de las frecuencias absolutas y relativas.

$X$	$n_i$	$N_i$	$f_i$	$F_i$
0	7	7	0.07	0.07
1	23	30	0.23	0.3
2	23	53	0.23	0.53
3	18	71	0.18	0.71
4	22	93	0.22	0.93
5	7	100	0.07	1

2. Completar la tabla marginal de frecuencias de la variable  $Y$ .

Análogo al apartado anterior, la tabla puede completarse del siguiente modo:

$Y$	$n_i$	$N_i$	$f_i$	$F_i$
0-10	20	20	0.2	0.2
10-20	30	50	0.3	0.5
20-30	30	80	0.3	0.8
30-40	20	100	0.2	1

3. El número máximo de asignaturas que tiene que haber aprobado un alumno para estar entre 73% de alumnos que menos asignaturas han aprobado.

Hay que tener en cuenta que el dato solicitado corresponde al percentil 73 de la variable  $X$ . Su valor es 4.

El número de horas de estudio a la semana que tiene que dedicar un alumno como mínimo para estar entre el 35% de los que más horas ha estudiado.

El dato solicitado corresponde al percentil 65 de la variable  $Y$ . Como la variable  $Y$  es continua el valor del percentil 65 viene dado por

$$L_{i-1} + \frac{\frac{65n}{100} - N_{i-1}}{n_i} a_i = 25$$

4. El número de asignaturas aprobadas más habitual es: (nota: en el caso de obtener más de un valor indicar sólo el más pequeño)

El número de asignaturas aprobadas más habitual es la moda de la variable  $X$  que corresponde a 1.

Teniendo en cuenta que la distribución de los datos es uniforme dentro de cada intervalo, el número de horas semanales dedicadas al estudio más frecuente es: (nota: en el caso de obtener más de un valor indicar sólo el más pequeño)

Como la variable  $Y$  es una variable continua, el primer paso consiste en identificar el intervalo modal (aquél que tiene una frecuencia absoluta mayor). Una vez identificado, la moda no es más que:

$$Mo = L_{i-1} + \frac{n_i + 1}{n_{i+1} + n_{i-1}} a_i = 16$$

Por tanto, el número de horas semanales dedicadas al estudio más frecuente es la moda de la variable  $X$  que corresponde a 16.

5. Para este apartado es necesario calcular el coeficiente de variación de cada variable,  $X$  e  $Y$ .

Dado que el coeficiente de variación de una variable viene dado por el cociente entre su desviación típica y su media:

$$CV_x = \frac{\sqrt{S_x^2}}{\bar{x}} \quad CV_y = \frac{\sqrt{S_y^2}}{\bar{y}}$$

Se tiene que el coeficiente de variación de la variable  $X$  es 0.57 y el coeficiente de variación de la variable  $Y$  es 0.51. Nótese que la media y varianza de la variable  $X$  son 2.46 y 1.99, mientras que para la variable  $Y$  son 20.57 y 110.23.

Por tanto, la distribución de la variable  $Y$  es más homogénea que la de la variable  $X$  ya que el coeficiente de variación de la variable  $X$  es mayor que el coeficiente de variación de la variable  $Y$ .

6. Sabiendo la siguiente información de las variables  $X$  e  $Y$ ,

$$\sum_{i=1}^{100} x_i = 246 \quad \sum_{i=1}^{100} x_i^2 = 804$$

$$\sum_{i=1}^{100} y_i = 2057.2 \quad \sum_{i=1}^{100} y_i^2 = 53343.78$$

Sabiendo que la covarianza entre dos variables viene dada por:

$$S_{xy} = \frac{\sum xy}{n} - \bar{x}\bar{y} = 14.43$$

se tiene que el coeficiente de correlación lineal es:

$$\rho = \frac{S_{xy}}{S_x S_y} = 0.9743$$

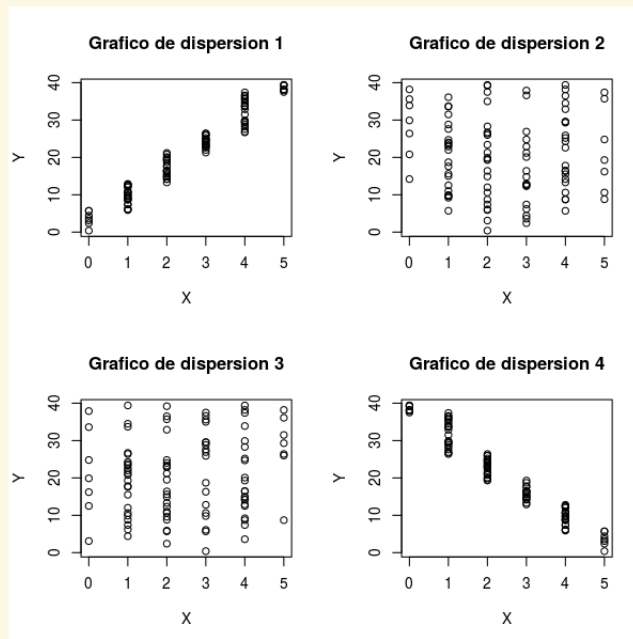
y el coeficiente de determinación lineal:

$$R^2 = \frac{S_{xy}^2}{S_x^2 S_y^2} = 0.95$$

y el coeficiente de determinación lineal:

$$R^2 = \frac{S_{xy}^2}{S_x^2 S_y^2} = 0.95$$

De las siguientes figuras:



La que mejor se ajusta a los resultados obtenidos es el diagrama de dispersión número 1 ya que el coeficiente de correlación lineal es cercano a 1 y esto indica una relación lineal entre las variables muy fuerte, es decir, la pendiente de la recta de regresión que ajustaría a los datos tiene pendiente positiva.

De todos los gráficos que se presentan, sólo el gráfico 1 se ajusta a ello.

7. El número esperado de asignaturas que aprobaría un alumno que estudia 32 horas a la semana viene dado por la recta de regresión:

$$\bar{x} + (32 - \bar{y}) \frac{S_{xy}}{S_y^2} = 3.96$$